DOCUMENT RESUME

ED 193 254 TM 800 568

AUTHOR Echternacht, Gary
TITLE Model C Is Feasible for ESEA Title I Evaluation.
PUB DATE Apr 80
NOTE 16p.: Paper presented at the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Admission Criteria; Analysis of Covariance; *Compensatory Education; Computer Programs; Elementary Secondary Education; *Models; Pretests Posttests; *Program Evaluation; School Districts; Statistical Analysis; Technical Assistance; Testing Programs
IDENTIFIERS Elementary Secondary Education Act Title I; *ESEA Models; Technical Assistance Centers; Title I Evaluation and Reporting System

ABSTRACT

The assertion that Model C is feasible for Elementary Secondary Education Act Title I evaluation, why it is feasible, and reasons why it is so seldom used are explained. Two assumptions must be made to use the special regression model. First, a strict cut-off must be used on the pretest to assign students to Title I and comparison groups. Second, the functional relationship between pretest and posttest must be assumed in order to fit a regression line for use in determining the performance level of students without Title I. A single composite measure that combines test scores, teacher ratings, and past academic performance can be used to select Title I students, and as the pretest. Model C needs only a district-wide assessment of all students and access to a computer to match the data. Six reasons given for not choosing Model C include: (1) Model C requires more work than does Model A; (2) local programmers lack expertise in manipulating scientific data; (3) large districts lack experience in matching pretest and posttest scores; (4) data analysis is necessary; (5) selection process must be open; and (6) normal curve equivalent gains are lower than when using Model A. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193254

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Model C is Feasible for ESEA Title I Evaluation

Gary Echternacht

Educational Testing Service

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G. Echternacht

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800568

Presented at

The American Educational Research Association's
Annual Meeting

April 10, 1980

OCT 1 2 1980

Model C is Feasible for ESEA Title I Evaluation

Gary Echternacht
Educational Testing Service

Background

Local school districts are required to use one of three different evaluation models to assess the effects of their ESEA Title I projects. Each of the three models makes this assessment in the same way--by estimating a level of student achievement that would have been observed at posttest time without Title I instruction and comparing that estimate with the observed posttest achievement. Comparisons for each model are expressed in terms of NCE gain (Tallmadge and Wood, 1976), and in the sense that all models provide NCE gains, the models produce comparable gains.

The model most commonly selected by school districts is the norm referenced model or Model A. Essentially, it is a gain score model and is the model held in lowest regard by statisticians and measurement people. Despite this disregard it is by far the most popular of the evaluation models. It is easy to use, requires the least amount of testing, requires little data analysis, and has an intuitive appeal that is easily grasped by the untrained.

Other than model A the model that is the most frequently used is called the special regression model or Model C. It is a variation of Belsen's (1956) method of stable correlates and the analysis of covariance (see Cochran, 1970). The model differs from the traditional analysis of covariance only in the way that the regression coefficient is estimated. In the Model C, the regression coefficient is estimated only in the

comparison group. In the analysis of covariance the regression coefficient is estimated from both the comparison and treated groups.

Two assumptions must be made to use the special regression model. First, a strict cut-off must be used on the pretest to assign students to Title I and comparison groups. A good statistician would say that we only need to know the probability that any given student with pretest score X is in the Title I group. Since program administrators do not like to assign students to compensatory programs based on probabilistic mechanisms, the probabilities must be one or zero, i.e., a student is either definitely in Title I if his pretest score is below a certain cut-off or definitely not in Title I if his pretest score is above the cut-off. If this assignment rule is violated, there will be some unknown amount of bias introduced in the estimate. Second, we must assume the functional relationship between pretest and posttest so that we can fit a regression line for use in determining the performance levels of students without Title I. Typically, one assumes a linear relationship between pretest and posttest although there is no reason why other relationships cannot be used (see Echternacht and Swinton, 1979). If we know the rule used to assign students to Title I and form of the regression line, Rubin (1977) has shown that unbiased results will be obtained from using the Model C analysis.

Several studies (e.g. Murray, 1978, Glass, 1978, and Echternacht, 1979) have shown that generally Model A results are biased. The analysis of covariance has been a tool often used by evaluators and statisticians to assess project effects in larger evaluations. The analysis of covariance is generally regarded as being the superior methodology, one that leads

to more accurate results than does Model A. Why then do so many districts choose Model A over Model C? Some would say that Model C is not practical and that despite its shortcomings Model A is the only feasible evaluation model that local districts can apply. This paper asserts that Model C is feasible, explains why it is feasible, and suggests reasons why Model C is so seldom used.

Model C is Feasible

From a theoretical standpoint, Model C requires only knowledge of the functional relationship between pretest and posttest and that a strict cut-off be used to select Title I participants. Title I program regulations require that the procedure used to select students be systematic and uniform for all students. A single composite measure that combines test scores, teacher ratings, and past academic performance can be used to select Title I students. Such a procedure is systematic, uniform, and meets all of the criteria found in Title I regulations. It is also perhaps the most comprehensive and valid procedure for identifying students most in need of the program. Although it requires more effort than districts usually give, it certainly is possible to construct such a composite and use the composite measure as the pretest.

It is a little more difficult to assume the functional form of the relationship between pretest and posttest. Most claims about the infeasibility of Model C come down to an assertion that we can never really know the form of this relationship. Tradition asserts that all relationships between tests are linear or very close to being linear. Most of educational evaluation and each of the other USOE models is based on

linear relationships between pretest and posttest scores. Our analyses, however, indicate that linear relationships cannot always be expected. When the relationship is non-linear and we use the usual Model C analysis with linear regressions, we introduce bias into the estimate, biases which are sometimes unacceptably large (see Echternacht, 1978). Nonetheless, as one gains experience with a district's data, patterns are recognized. When districts see non-linear relationships repeatedly, they will want to fit non-linear regressions. The technology is not that difficult. Only the experience in determining what type of relationship to expect is lacking.

From an administrative standpoint, Model C needs only a district-wide assessment of all students and access to a computer to match the data and crunch the numbers. District-wide testing programs abound. Even in Model A designs where only Title I students are tested, the posttest is often the test score obtained by students in the district-wide testing.

Although there are procedures that can be used to grind out the calculations on a hand calculator, there usually are so many analyses (by grade and subject matter) that only by keypunching and analyzing the data by computer does the procedure become feasible. This administrative need prevents some smaller districts from using Model C. If the district and project are small enough, however, it does become possible to use a hand calculator although in such instances the analysts generally lack the statistical background to perform the analysis.

As an example that Model C is feasible, let me describe the work that our technical assistance center has been doing with the Bayonne NJ school district. Bayonne chose Model C because they could use the model without any additional testing. For the most part, students were selected for the program based on the spring district-wide test results. The district had also contracted with the computer science department at nearby St. Peters College to prepare Title I evaluation and management reports. Our center has been providing assistance to Bayonne for three years now. The Title I staff there has minimal training in statistical analysis.

During the first year, Model C was used in only two grades. St. Peters performed the analysis for the district, using flow-charts prepared by our center. Matching pretests and posttests were the most costly and difficult aspect of the analysis. St. Peters has subsequently worked with the district on means of matching the data, and today I am happy to report, the matching is carried out efficiently.

Our center verified the results for the first year and showed Title I staff at Bayonne how to interpret the results to nonstatistical audiences. The analysis differed from the usual Model C analysis in that building level estimates were obtained using two different methods, then building estimates were aggregated to produce the district NCE gain at the grade level. The analysis methods differed in the way the regression coefficients were estimated. Individual building level estimates were used in one method; pooled estimates were used in the other method.

In the second year, the analysis was expanded to all grades. At this time the selection issue came into play. Strict cut-offs had not been observed in the past. The district chose to modify their selection procedures so that a strict cut-off was closely adhered to. They did not feel threatened by this move, and indeed were pleased when they were able to demonstrate clearly to State Title I monitors exactly how selection was conducted. The move toward more objective selection was viewed as a positive side benefit of using Model C.

In the third year, the computer program was again modified at the suggestion of our center. Two additional analyses--the analysis of covariance with building level and pooled regression coefficients--were added. District staff feel comfortable with the analysis although they must depend on the technical assistance center for advice when certain problems arise. Using the computer center at St. Peters has been a move that other districts might imitate. Their advice and assistance with regard to data processing and reporting has been excellent.

Why is Model C so seldom used?

If Model C is feasible, why don't more districts use it? Since many districts have access to data processing services and want a systematic student selection system, why do they so frequently choose not to use Model C? Our experience in training and consulting with school districts indicates there are six major reasons for not choosing the model. They are:

1. Model C requires more work than does Model A.

2. Local district computer programmers have little or no experience in manipulating scientific data.
3. Local districts have little experience in matching pretest and posttest scores for large numbers of students.
4. Evaluators using Model C must have some skill in data analysis.
5. Model C requires that the selection process be open for all to see.
6. Model A results in higher NCE gains than when using Model C.

Since Model C requires that data be collected on non-Title I students as well as on Title I students, the total volume of data processed is much larger than it is for Model A. Traditially, local evaluators have not collected data on non-Title I students, and when they choose Model C, they are choosing to process sometimes as much as two to ten times as much data as they are accustomed to. Since the data collection burden often falls on one person, this increase in the amount of data collected and processed is one reason why school districts do not use Model C.

Most large and many small school districts either have or have access to a computer. Computers are widely used to process payroll, schedule, and perform management functions. Most computer programmers in school districts use COBOL as a programming language and find it difficult to use matricies, arrays, and subroutines that form the heart and soul of scientific work. Model C computer programs are written in FORTRAN, and there is a reluctance for COBOL programmers to become familiar with scientific programs.

Title I teachers usually have been responsible for matching up pretest and posttest scores for their students. They have paid little attention to non-Title I students. No one in local school districts seems to be responsible for matching scores for the other students. In Model C the matching is often performed by computer according to some identification number. Most school districts have not established permanent student identification numbers, and for matching they must rely on birth dates or scoring service assigned identifiers, both of which are likely to result in either coding or administrative errors on the part of students or teachers.

In applying Model C, the evaluator must act as a data analyst. The analysis must be conducted in stages where a decision is made in each stage about the approach used in the next. A good analyst would want to look over plots of pretest and posttest scores before fitting a regression function. Plots lead the analyst to further investigations in an attempt to explain the resulting NCE gains achieved. Local evaluators typically have not approached evaluation in this manner. Evaluators often have responsibilities to other programs. They have a short time to do their anlayses. Evaluators have applied their analyses, reported the results, and have done little in the way of explaining why the results turned out as they did (see David, 1978). In the training that we give local districts, we emphasize the need for acquiring some data analysis skills, and often they reject the model because they feel that they cannot learn such skills.

Although Title I regulations require that the selection for the program be on the basis of systematic and uniform criteria, school districts often assign some students to Title I on arbitrary grounds. Some students are assigned on the basis of teacher or counselor insistence even though test scores and other indicators of ability show that the student is not in the "neediest" category. Students who are behavior problems often wind up in Title I. If districts use Model C then the selection criteria are explicit and easy to monitor. Districts fear such monitoring and often choose model A which does not specify how students are selected. Some school administrators are quite frank about not choosing Model C for this reason. They indicate clearly that they would rather be found "in compliance" with respect to student selection and have a weak evaluation design than they would be found "not in compliance" with respect to student selection and have a strong evaluation design.

Finally, in cases where both Model A and Model C have been applied (Hardy, 1979) comparison of the resulting NCE gains has led Model C districts to change to Model A. With the exception of RMC, who developed the system, most investigators into the issue of the comparability of the models have concluded that the three models produce different NCE gains. Experience from the field has indicated that in general Model C results are lower than Model A results. In Florida results from the two models were compared. Model C estimates of effect were much lower than Model A estimates, and districts who had been using Model C began to switch to Model A. Most districts like to show their projects in a favorable light, as it reflects on the abilities of administrators and teachers alike. Using Model A is one way to show a project at its best.

Making it easier to use Model C

State departments of education and the federal government can do a number of things that would make it easier for local districts to use Model C. A conscious effort must be made to demonstrate that it is in the school district's best interest to have a strong evaluation design, and that evaluation models should provide long term indicators of the Title I programs's effectiveness. Title I program. Evaluation must be viewed as a continuous process that leads to observable improvements in services. School district people are genuinely interested in doing the extra work required in Model C if they see a tangible payoff. District people must be convinced of the need for accurate project impact estimates and shown how to use those estimates as a trigger for changing the educational process.

State departments can also make more use of the regional technical assistance centers in providing training in data analysis and data management techniques. Our experience with Bayonne indicates that much training and communication needs to take place before a district can go on its own with the model. Periodic visits over a number of years are required. A rapport between the district staff and technical assistance center staff must develop so that issues can be resolved and problems investigated. Because technical assistance centers have limited resources, by necessity the number of Model C sites must be somewhat restricted. But district staff can learn to match pretest and posttest scores, they can be provided with computer programs, and they can learn sufficient data analysis skills to at least identify problems and refer them to the

technical assistance center. Identifying a few districts as Model C sites and targetting technical assistance center services to those sites provide districts with the necessary training and experience.

To overcome the problem with the strict cut-off and selection in general, state departments should emphasize adherence to the Title I regulations regarding student selection. There is also a training need here. Often when district people make what turns out to be arbitrary assignments to Title I, they do so because they believe that the person arbitrarily assigned is in most need. When they do this, they are only failing to note the specific area of need so that it can be assessed uniformly for all students. District staff are not always the villains when found to be not in compliance with respect to student selection. Often they only lack the knowledge of how to translate valid criteria into objective measures.

If the notion that the models produce "comparable" results could be reformulated, districts might not search for the methodology that gives the largest NCE gains. The word "comparable" is not a good word. In the legislation creating the Title I evaluation system, the Commissioner is required to develop "comparable models" for school districts to use in evaluating their Title I programs. But in what sense should the models be comparable? When RMC developed the models, they took the word "comparable" to mean will give identical results. In this sense all evaluation results might be averaged to produce a single statistic indicative of the success of Title I nationally at a given grade. Of course this type of thinking is naive. Every good statistician knows

that different analyses will give different results. The results might lead to the same conclusion, but the specific details--the estimates--will differ. Every good lawyer knows that evidence from multiple sources is more convincing than is evidence from a single source. So, too, should it be with the Title I evaluation and reporting system. Model A and Model C results should not be compared against each other. They are two different indicators of overall effectiveness, with one methodology being the more rigorous statistically, the other being the only methodology possible at the sites. The national impact of Title I could then be represented as sets of estimates with a conclusion based on the whole set of results rather than a single indicator.

Reference Notes

Glass, G.V. Memorandum to technical assistance center project directors. 1978

Hardy, R. A comparison of Model A and Model C: results of first year implementation in Florida. Discussion paper for TAC directors meeting, January, 1979.

Reference

Belsen, W. A technique for studying the effects of a television broadcast. Applied Statistics, 1956, 5, 195-202.

Cochran, W. The use of covariance in observational studies. Applied Statistics, 1970, 18, 270-275.

David, J. Local uses of Title I evaluations. Research report EPRC 21. Menlo Park, CA: SRI International, July 1978.

Echternacht, G. The use of different models in the ESEA Title I evaluation system. Paper presented at the AEA annual meeting, Toronto, 1978.

Echternacht, G. The comparability of different methodologies for ESEA Title I Evaluation. Paper presented at the APA annual meeting, New York, 1979.

Echternacht, G., and Swinton, S. Going Straight. Paper presented at the AERA annual meeting. San Francisco, 1979.

Murray, S. An analysis of regression effects and the equipercentile growth assumption in the norm referenced evaluation model. Paper presented at the Washington Educational Research Association annual meeting, Seattle, 1978.

Rubin, D. Assignment to treatment group on the basis of a covariate. Journal of Educational Statistics, 1977, 2 (1), 1-26.

Tallmadge, G. K. and Wood, C. User's Guide: ESEA Title I evaluation and reporting system. RMC Corporation: Mountain View, CA, 1976.